Simple Linear Regression

Adv. expenditure (X) vs Sales (Y)

└ dependent var

└ independent var

Model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

└ Error term

└ Independent var (Adv. \$)

└ slope

└ Y intercept

└ dependent (Sales \$)

→ use SMA, WMA or EXP

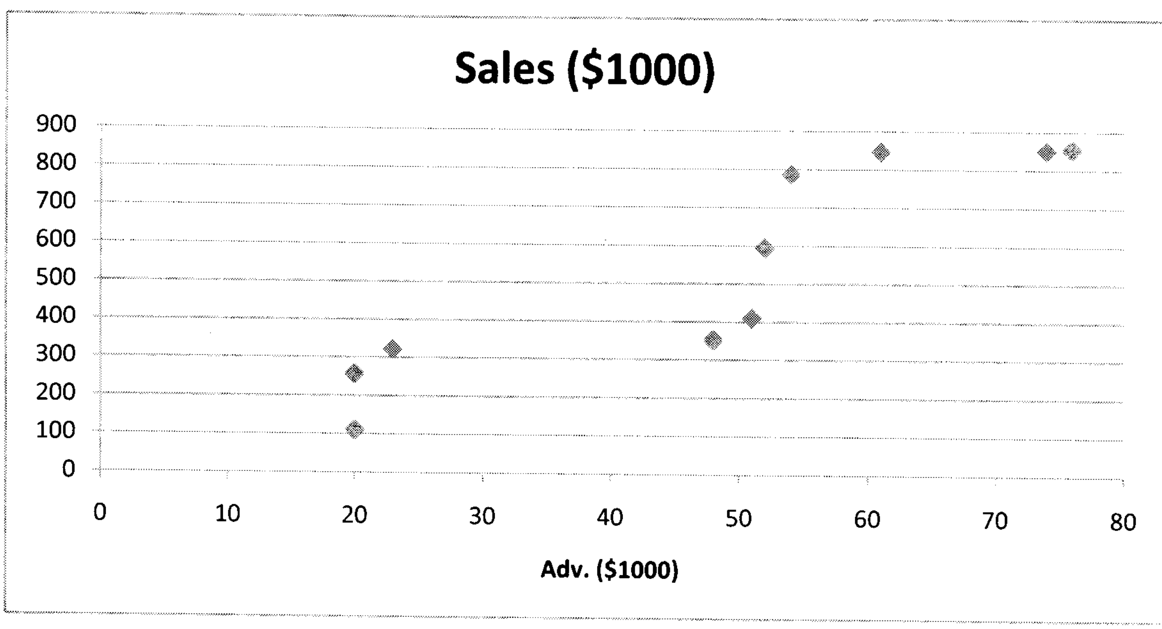
Seasonality

trend → Regression

Linear Reg

Scatter plot (stationary vs. Nonstationary)

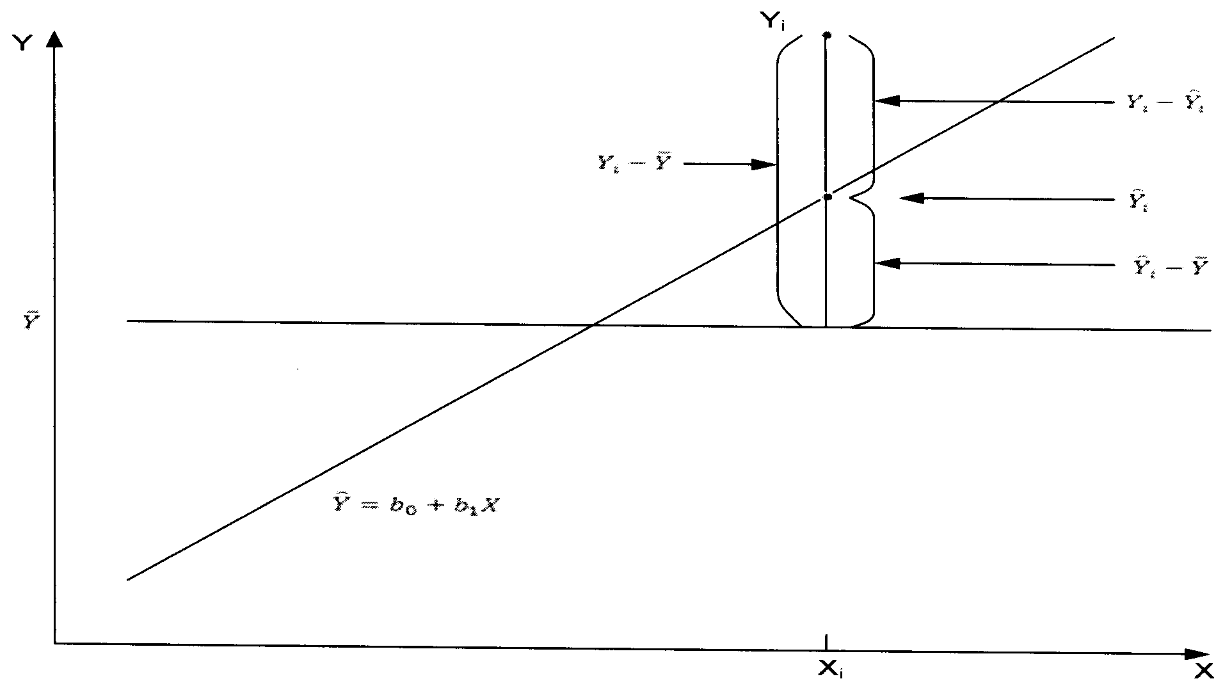
Adv ($1000)	Sales ($1000)
20	112
20	259
23	323
48	353
51	411
52	594
54	788
61	846
74	851
76	854
35	
45	



Fit a straight line with
- any two points
- Run and Rise plus y intercept
- Least Squared Errors (LSE)

3. Using Excel@ to do Regression analysis

$\hat{Y} = b_0 + b_1X$ $\bar{Y}$ • X_i $Y_i - \bar{Y}$ $Y_i - \hat{Y}_i$ $\hat{Y}_i$ $\hat{Y}_i - \bar{Y}$



Decomposition of the Total Error: $Y_i - \bar{Y} = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y})$

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$$

TSS = ESS + RSS

Or Total Sum of Squared Errors (TSS) = Error Sum of Squares (ESS) + Regression Sum of Squares (RSS)

$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}$, and $0 \le R^2 \le 1$

R^2 refers to the percentage of the total variation of Y around its mean that is explained or counted for by the estimated regression line or how well the regression line fits the data.

$1-R^2$ is the percentage of the total variation of Y around its mean that is unexplained or uncounted for by the regression line.

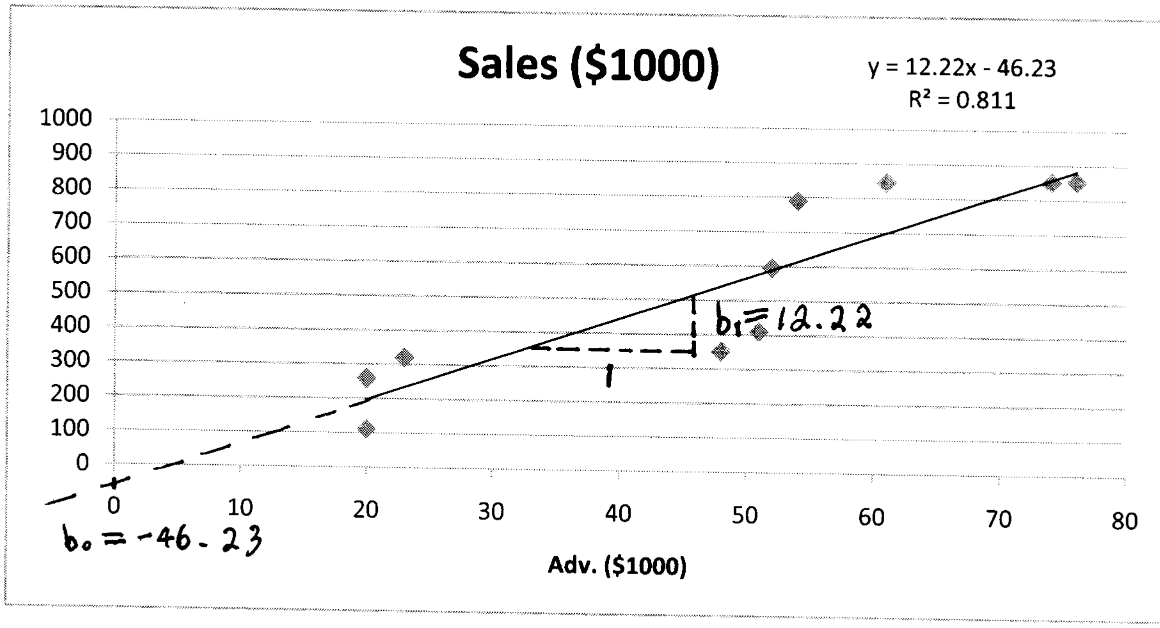
Equations to compute b_0 and b_1:

$$b_1 = \frac{SSXY}{SSX} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n} X_iY_i - \frac{[\sum_{i=1}^{n} X_i][\sum_{i=1}^{n} Y_i]}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{[\sum_{i=1}^{n} X_i]^2}{n}}$$

$$b_0 = \bar{Y} - b_1\bar{X}$$

Click any data point on the scatter plots for Month and Sales, or Adv and Sales, select Add Trendline / Display equations & Display R-Squared value on the charts. The Y and Xs are renamed to Month, Adv and Sales, respectively, for the regression lines.

Adv ($1000)	Sales ($1000)
20	112
20	259
23	323
48	353
51	411
52	594
54	788
61	846
74	851
76	854
35	
45	



Use Excel@ Data/Data Analysis/Regression to get the Summary Output for the data and print a copy of it, find values of b_0, b_1, and b_2 in the Summary Output. The values of b_0, b_1, and b_2 are labeled in the Summary Output below.

Meanings of Regression Summary Output

R^2

Adjusted R^2

S_e

n

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.981
R Square	0.963
Adjusted R Square	0.953
Standard Error	61.174
Observations	10

SSR SSE SST MSR MSE p-value Regression

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	688012.852	344006.426	91.924	9.45011E-06
Residual	7	26196.048	3742.293		
Total	9	714208.9			

Confidence Intervals for b_0, b_1 and b_2

b_0

b_1

b_2

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	120.743	59.583	2.026	0.082	-20.147	261.633
Month	135.828	25.262	5.377	0.001	76.091	195.564
Adv ($1000)	-6.862	3.682	-1.864	0.105	-15.569	1.845

S_{b0} S_{b1} S_{b2} p-value to test $b_0 = 0$ p-value to test $b_1 = 0$

p-value to test $b_2 = 0$

(Regression.xls/Reg1SOa)

General Linear Regression Model:

$$y_i\,(Sales) = \beta_0 + \beta_1 X_{1i}(Month) + \beta_2 X_{2i}(Adv.) + \varepsilon_i$$

β_2 – amount of increase of sales when X_{2i} increase by 1

β_1 – amount of increase of sales when X_{1i} increase by 1 month

β_0 – amount of sales when $X_{1i} = \emptyset$ & $X_{2i} = \emptyset$

Estimated Linear Regression Function (Equation)

$$\hat{y}_i\,(Sales) = b_0 + b_1 X_{1i}\,(Month) + b_2 X_{2i}\,(Adv.)$$
$$= 120.74 + 135.83\, X_{1i} - 6.86\, X_{2i}$$

-6.86 – Sales dereased by -6.86 when Adv. increased by 1.

135.83 – Sales increased by 135.83 when month increased by 1 month.

120.74 – Sales is 120.74 when both Month $= \emptyset$ & Adv $= \emptyset$

Forecast Sales when Month (X_1) $= 11$ and Adv. (X_2) $= 2$

$$\hat{Y}\,(Sales) = 120.74 + 135.83 * 11 - 6.86 * 2 = \$1469.15$$

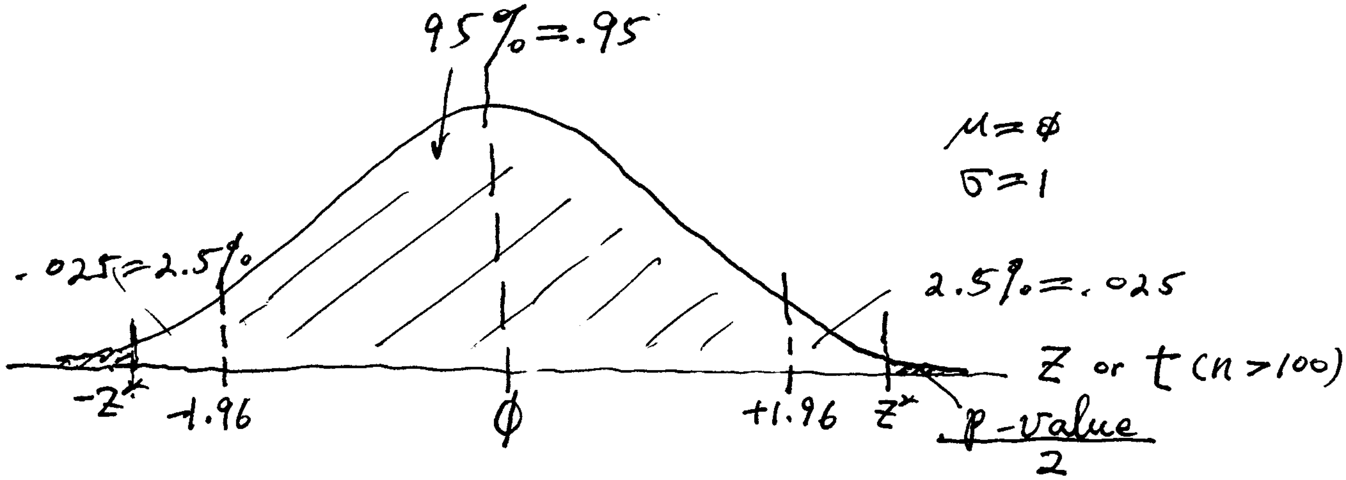
Standard Error $= S_e = S_{yx} = 61.174$

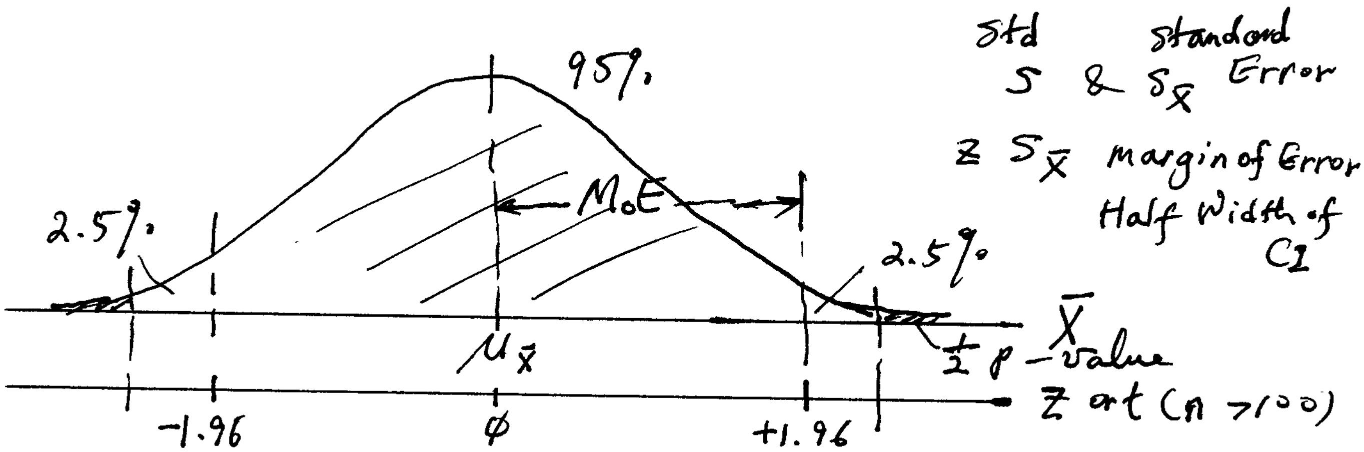
Approximated 95% CI for prediction of Y given X value (month = 11 & adv. = 2)

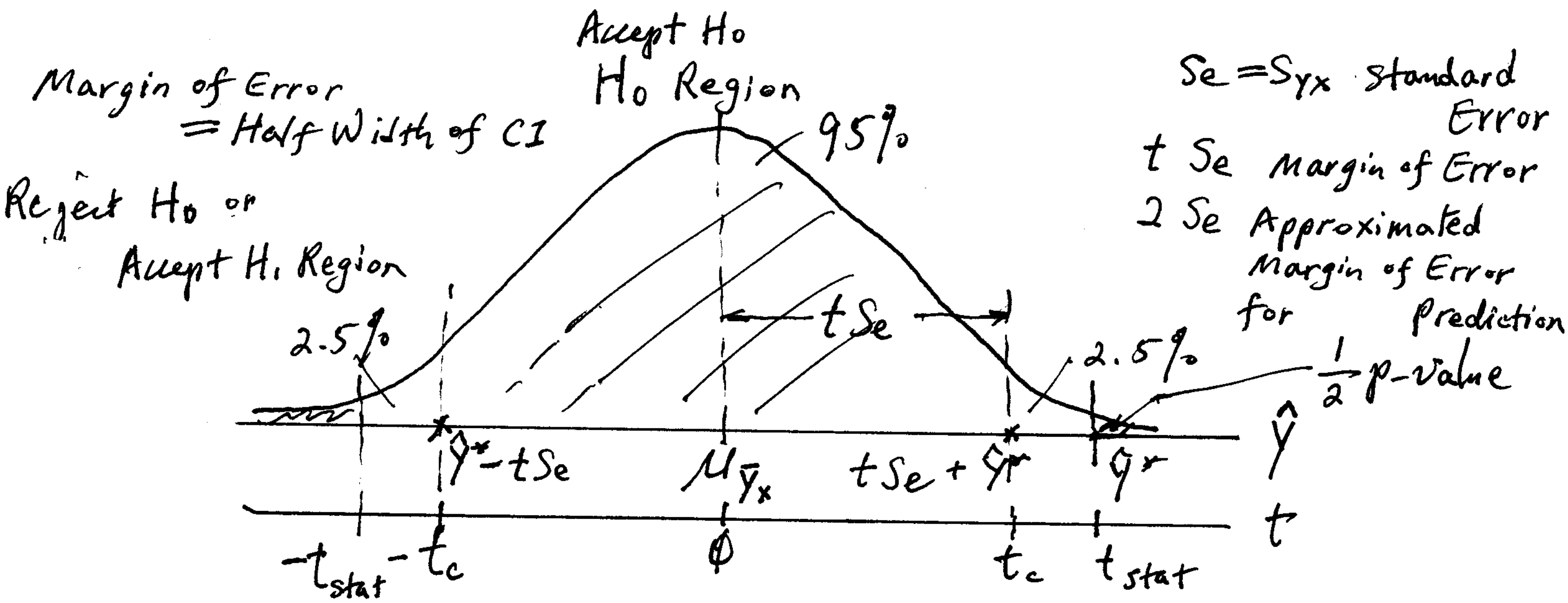
Low Limit: $\hat{y} - 2S_e = 1469.15 - 2 * 61.174 = \$1346.80$

Upper Limit: $\hat{y} + 2S_e = 1469.15 + 2 * 61.174 = \$1591.50$

We are 95% sure the unknown true sales, when month = 11 and adv. = 2, is approximately between 1346.80 and 1591.50.


95% = .95
μ = 0
σ = 1
.025 = 2.5%
2.5% = .025
Z or t (n > 100)
-Z* -1.96 0 +1.96 Z*
p-value / 2


Std S & S X̄ Standard Error
z S X̄ margin of Error
Half Width of CI
95%
MoE
2.5%
2.5%
X̄
½ p-value
μ X̄
Z or t (n > 100)
-1.96 0 +1.96


Accept Ho
Ho Region
Margin of Error
= Half Width of CI
Reject Ho or
Accept H1 Region
Se = Syx Standard Error
t Se Margin of Error
2 Se Approximated Margin of Error for Prediction
95%
t Se
2.5%
2.5%
½ p-value
ŷ* - tSe
μ ȳx
tSe + ŷ*
ŷ*
ŷ
t
-tstat -tc 0 tc tstat